HHS Public Access

Author manuscript

Curr Opin Struct Biol. Author manuscript; available in PMC 2017 October 13.

Published in final edited form as:

Curr Opin Struct Biol. 2016 October ; 40: 145–152. doi:10.1016/j.sbi.2016.09.013.

Multi-level regulation of cellular glycosylation: From genes to transcript to enzyme to structure

Sriram Neelamegham[1] and **Lara K. Mahal**[2]

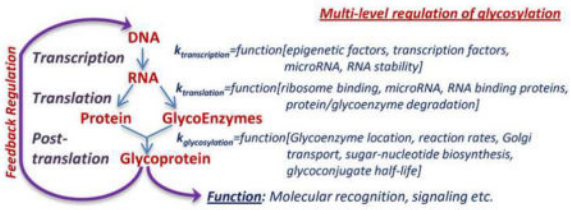
[1]Chemical and Biological Engineering, and Clinical & Translational Research Center, State University of New York, Buffalo, NY 14260, USA

[2]Department of Chemistry, Biomedical Chemistry Institute, New York University, New York, NY 10003, USA

Abstract

Glycosylation is a ubiquitous mammalian post-translational modification that both decorates a majority of expressed proteins and regulates their function. Cellular glycan biosynthesis is facilitated by a few hundred enzymes that are collectively termed ‘glycoenzymes’. The expression and activity of these enzymes is controlled at the transcription, translation and post-translation levels. New wet-lab advances are providing analytical methods to collect large-scale data at these multiple levels, relational databases are starting to collate these results, and computer models are beginning to integrate this information across scales in order to gain new knowledge. These activities are likely to enable the qualitative and quantitative mapping of pathways regulating glycan production and function in proteins, cells and tissue.

Graphical Abstract



Introduction

Cell surface, secreted and intracellular proteins and lipids are extensively glycosylated. These modifications affect a range of molecular recognition and signaling events [1]. Whereas intracellular glycoproteins often contain short carbohydrate structures, commonly bearing a single O-linked GlcNAc (N-Acetyl glucosamine) residue, extracellular glycans

Address correspondence to: Sriram Neelamegham, 906 Furnas Hall, State University of New York, Buffalo, NY 14260, USA; Tel: 1-716-645-1200; Fax: 1-716-645-3822; neel@buffalo.edu.

bear more elaborate structures [2]. On human cells, these include branched O- and N-linked glycans, glycolipids and repeating glycosaminoglycans (Figure 1). These sugars are commonly made up of nine different monosaccharides. Altogether, human glycoproteins and glycolipids are estimated to contain a minimum of 3,000 unique 2–5 monosaccharide determinants [3]. Individual cell and tissue only have a subset of the glycoenzymes and thus individual glycome typically have fewer members. Similarly, glycosaminoglycans are estimated to contain up to ~4,000 possible pentasaccharide sequences [3]. These determinants provide the recognition motifs underlying the biological functions of glycans.

Classical studies into the biosynthesis of glycan epitopes use reductionist, biochemical methods to elaborate individual biosynthetic steps. In recent years, however, newer ‘omics’ approaches are rapidly bridging gaps in our knowledge of glycan biosynthesis, especially with respect to cell-type and context-specific glycosylation [4–7]. These newer experimental modalities often use mass spectrometry and lectin microarray based technologies [8]. Integrating these glycomics data sets with high-throughput genomics, transcriptomics, enzymology and proteomics methods is necessary as it can lead to a ‘systems level’ understanding of glycan biosynthetic pathways [9]. This review describes progress and challenges in the development of such systems-based understanding, with emphasis on computer based modeling and pathways regulating canonical O- and N-linked glycosylation in humans.

Multi-level regulation of glycosylation

The regulation of glycoproteoforms (i.e. the glycosylation variants of a protein) in metazoans is complex. It involves not only the biosynthesis of the underlying protein scaffold built on its mRNA template, but also non-template driven post-translational modifications that is driven by the glycosylation-related enzymes or ‘glycoenzymes’. In humans, ~500 glycoenzymes participate in this process including 209 glycosyltransferases, 76 glycosidases, 114 enzymes involved in sugar metabolism & transport, 54 sulfation-related enzymes and 31 enzymes regulating additional lipid and GPI-anchor biosynthesis pathways [10,11]. Regulation of protein scaffolds and glycoenzymes (themselves proteins) underlying glycoproteoforms occurs at multiple steps, including transcriptional, translational and post-translational levels (Figure 2).

Transcription is perhaps the most recognized regulator of protein, and thus glycoproteoform, expression. Transcription of a glycosylation-related gene (glycogene) is controlled by both transcription factors binding to gene promoter and enhancer elements, and by epigenetic regulation of gene accessibility [12]. Transcription factors, which can act as promoters or repressors, are often glycosylated by O-GlcNAc which can regulate the transcriptional machinery [2]. In general, there is little available information about transcription factor-mediated control of glycogene expression, with only ~2 dozen glycogene-transcription factor interactions known to date and mechanistic work available for ~15 glycogenes [13–16]. The importance of transcription factors in control of glycosylation is exemplified by recent work on the transcription factor XBP1s, a major regulator of the unfolded protein response (UPR) [15]. Here, the artificial activation of XBP1s induced expression of a series of *N*-glycan biosynthetic enzymes, shifting the glycosylation profile of a secreted

glycoprotein from oligomannose towards more processed hybrid and complex epitopes, tying together the UPR with changes in the glycome [15]. In another example, the phosphorylated form of the transcription factor ATF2, which has a role in therapeutic resistant melanoma, was recently found to directly repress expression of fucokinase (FUK). This enzyme is a key component of the fucose salvage pathway and its repression by activated ATF2 resulted in ~40% loss of cellular fucosylation [16]. Beyond transcription factors, access to the genome, which is controlled by epigenetic modifications such as acetylation, methylation and ubiquitinylation also regulate glycogene transcription [12]. This is even less studied than transcription factor binding to glycogenes, although a recent study analyzing DNA methylation of 86 glycogenes suggests a positive correlation between the promoter methylation status of glycogenes and expected tumor-associated glycan structures [17]. A handful of studies have also shown a direct interaction between the methylation status of specific glycogenes and glycogene expression [18–21]. In this regard, the methylation of the α(2,3)sialyltransferase ST6Gal1 promoter was shown to silence expression of the ST6Gal1 transcript in bladder cancer [18], confirming an earlier report in breast cancer showing methylation-dependent silencing of this glycogene [22]. Mapping transcription factors and epigenetic regulation onto glycogene regulatory networks will require a combination of methods from traditional Chromatin Immunoprecipitation (ChIP) assays to more sophisticated promoter-reporter assays that can reveal indirect and direct effects, to more sophisticated analysis of genome accessibility and silencing [15,23]. All of this is important to our understanding of the glycome and its regulation.

Although transcript (i.e. cellular mRNA) levels are often used as a key indicator of what glycogenes are being translated and the levels of biosynthetic enzymes and glycans that result, the relationship is not straightforward. A number of investigators have attempted to relate transcript levels to glycoenzyme activity and carbohydrate structures, including the response of the system to perturbations [11,24–26]. Such studies have been conducted using human promyelocytic leukocytes differentiated to primary neutrophils [24], a limited panel of mouse tissue [11], mouse embryonic stem cells differentiated to embryoid bodies and endodermal cells [25], and epithelial ovarian cancer cell types [26]. The broad conclusion of these studies is that glycogene, and more specifically glycosyltransferase, expression measurements in a number of cases can semi-quantitatively predict corresponding enzyme activity and cell-surface glycan expression. However, the relationship is often non-linear and in some instances, changes in glycan structures upon perturbation do not correlate with changes in glycogene mRNA expression levels. Although part of this discrepancy may be due to difficulties with quantitatively measuring low-abundance glycogene transcript levels [11], the post-transcriptional regulation of glycogene expression may also control carbohydrate structures.

MicroRNAs (miRNAs), small non-coding RNAs, are major post-transcriptional regulators of protein expression, acting as rheostats to maintain key proteins critical to a cell state at set levels [27,28]. Genes with lower expression levels, such as glycogenes, are both preferentially targeted by miRNA and more impacted by miRNA binding [28]. Recent work has shown that miRNA are indeed a major regulator of the human glycome and the impact of miRNA on the glycome via alteration of glycosyltransferase protein levels may explain discrepancies with the transcriptome [4,29]. For example, miR-34a, a microRNA associated

with hepatocellular carcinoma, targets FUT8. This miRNA did no impact FUT8 transcripts in human HepG2 cells, although it concomitantly altered FUT8 protein expression and core fucosylation levels [30]. miRNA regulate biological pathways by regulating sets of gene transcripts. Multiple laboratories have demonstrated that downregulation of single glycogene (or other gene) targets of miRNA mimic the biological effects of the miRNA [31–33]. Thus, miRNAs could be used as a proxy to identify glycosylation enzymes involved in the pathways they regulate (miRNA proxy approach) [32,33]. The miRNA proxy approach has been applied to identify functionally relevant glycogenes impacting cell states like epithelial-to-mesenchymal transition (EMT) [32] and the cell cycle [33]. For example, miR-200b, a regulator of EMT, was found to regulate a set of glycogenes from hard to analyze cohorts those involved in glycolipid (ST3GAL5 and ST6GALNAC5) and non-canonical O-glycan biosynthesis (B3GLCT). Downregulation of these glycogenes using shRNA phenocopied the miRNA, showing a role for these genes in the control of EMT [32]. This work opens up the possibility of leveraging post-transcriptional regulation of glycosylation as a means to reveal critical functional changes in the glycome that underlie biological processes. This may be especially important for the portions of the glycome, such as glycosylaminoglycans and glycolipids, which are overlooked in most analysis due to analytical challenges.

In addition to transcription, the measurement of glycoenzyme translation, turnover rates and activity are key parameters controlling glycan biosynthesis. In this regard, newer technologies like ribosome profiling are enabling a more nuanced understanding of protein synthesis [31]. This technology uses deep sequencing to identify stretches of mRNA that are actively bound to the ribosome during *in vivo* translation. Such quantitative estimation of protein biosynthesis rates may provide a new avenue for evaluating translation, not just transcription, as a measure of glycoenzyme expression, taking into account posttranscriptional regulation by miRNA.

Beyond expression, the direct measurement of glycoenzyme activity *in situ* is important. However, currently there is no established method for this measurement in the milieu of the Golgi. In complex mixtures, enzyme activity is typically followed in reaction vessels by assaying the conversion of a synthetic acceptor to product using cell/tissue lysates as the enzyme source. Product quantitation methods range from radioactivity based chromatography assays [34], to fluorescence based strategies that quantify glycan biosynthesis in an array format [35,36] and mass spectrometry [37]. Here, solution based enzymology assays are more sensitive compared to studies performed in 2D-array format, since only limited amounts of the carbohydrate acceptor can be immobilized on an array spot. As an alternative, mass spectrometry can be used to follow glycan biosynthesis in cells by feeding them small molecule acceptors, such as per-acetylated Benzyl-α-GalNAc, and subsequently measuring carbohydrate products formed on the synthetic substrate [38]. Data from such experiments, in tandem with genome editing methods (TALEN, CRISPR-Cas9 etc.), may enable estimation of glycosyltransferase activities and dissection of the interdependencies among glycogenes in the same biosynthetic pathway [5,39].

Overall, high-throughput analysis of the transcriptome is starting to provide clues to the nature of the cellular glycome and its response to system perturbation. Integration of

additional data including miRNA, glycoenzyme expression and activity is necessary in order to obtain a more complete picture of the glycan regulatory process.

Integrating knowledge using computer models of glycosylation reaction networks

The analysis of large-scale experimental datasets available from the above technologies may be streamlined by developing computational tools that can integrate results collected using these diverse methods. While such computational tools have been widely developed for genomics and proteomics research, they lag behind in the glycosciences [9,40]. This is in part due to the greater complexity of this research field, since glycoconjugate biosynthesis is regulated by multiple pathways including: i. sugar-nucleotide donor biosynthesis (e.g. UDP-Galactose) in the cellular cytoplasmic and nuclear compartments; ii. glycosylation reactions in the endoplasmic reticulum (ER) and Golgi cisternae; iii. transport processes that result in the sub-cellular localization of glycoenzymes, glycoconjugates and other reaction components in specific compartments; iv. trafficking of glycoproteins and lipids through the ER and Golgi which is impacted by cellular lectins and other transport mechanisms; and v. salvage pathways that aid the recycling of glycoconjugates back to their basic building blocks (Figure 2).

Integration of data across scales requires a robust modeling framework to formulate glycosylation reaction networks, integrate data from different relational databases, and enable code sharing among laboratories. Due to the last requirement, the Systems Biology Markup Language provides the ideal schema for model building [41,42]. Although this modeling language does not have specific facilities to handle glycan structures and glycosylation related reaction pathways, the annotation fields of the eXtensible Markup Language (XML) files can be easily modified to store pertinent information. For example, the annotation element of the *Species* field could store glycan structure information in XML format, the *Reaction* field can hold glycoenzyme and carbohydrate transport definitions/data, and *Compartment* is available to describe the physical features and reaction mechanisms in the ER and Golgi cisternae [42,43].

Some initial examples of glycosylation pathway modeling have appeared, and these investigations suggest that quantitative analysis of glycan structure data using computer modeling can yield novel biological insight. Notably, the computer models of Lau *et al.* showed that the activities of various N-acetylglucosaminyltransferase enzymes (MGAT1–5) together with the intracellular concentrations of UDP-GlcNAc regulate the pattern of N-glycan branching on cell surface proteins, especially growth factor receptors [44]. Higher levels of N-glycan-branching results in more robust receptor cross-linking (or lattice formation) using multivalent galectin-3, longer receptor retention/lifetimes on the cell-surface, and consequently more pronounced cell signaling. In another study of O-linked glycosylation, Liu *et al.* studied the critical enzymes (fucosyltransferases and sialyltransferases) regulating microheterogeneity in the leukocyte cell surface protein P-selectin glycoprotein ligand-1 and leukocyte cell adhesion function [45]. Predictions from such modeling have been validated by developing RNA-interference and CRISPR-Cas9

based knockouts [39,46]. In another example, Bennun *et al*. modeled cancer cell transcriptome and glycome data to demonstrate augmented α(1,2)fucosylation on prostate cancer cells during their transition from androgen-dependent to androgen-independent lymph node carcinoma [47]. Besides these studies that are of biomedical relevance, several studies have modeled the glycosylation status of recombinant proteins produced using CHO cell lines as this is of biopharmaceutical relevance (reviewed in [48]).

While some examples of glycosylation networks have appeared as illustrated above, additional development of machine-readable glycoenzyme definitions can accelerate such modeling efforts as this will enable: i) model-sharing among investigators, and ii) streamlined *in silico* description of different biological aspects such as sugar-nucleotide biosynthesis, species transport and rate expressions. In this regard, XML-based definitions for glycosyltransferases and glycosidases have appeared for O- and N-glycosylation pathways [42,43]. Succinct, computationally efficient definitions for the specificity of enzymes involved in N-linked glycosylation have also been described, though these definitions appear in string rather than XML format [47]. McDonald *et al*. also describe a similar modeling language specifically for O-glycosylation [49]. It seems possible that the merits of the above approaches may be combined to develop a more universally acceptable and comprehensive glycoenzyme database. Besides being linked to the GenBank for gene-level information and the carbohydrate-enzyme database CAZY for glycoenzyme structure and carbohydrate-binding module information [50], such a resource may also be parameterized using enzyme catalytic activity and rate constant data stored in BRENDA. Additionally, the glycoenzyme definitions may include reaction descriptions from the IUBMB ExploreEnz database, and pathway data from KEGG GLYCAN, an extension of the Kyoto Encylopedia of Genes and Genomes (KEGG) [51] (Table 1). Such efforts can enable the seamless integration of computational models with a variety of glycosylation related databases.

In addition to model formulation, advanced computer simulation methods are also needed in the field. In this regard, glycosylation reaction networks typically have a large number of reactions and reactants, but relatively few model parameters due to the limited number of glycoenzymes in a given system. Fitting such over-determined systems using deterministic ordinary differential equation (ODE) networks is complicated, though this is commonly done [44,45,47,52]. In such studies, optimization approaches like genetic algorithms are sometimes used to determine the *bounds* of the enzyme rate constants rather than exact solutions [45]. Another approach uses stochastic Markov chain methods for data fitting [53]. This method is computationally efficient but empirical (i.e. not mechanism based) in that similar reactions catalyzed by the same enzyme may have different rate constants. Besides these approaches, there are few examples of other computational techniques in the field like Boolean networks, agent-based models and statistical modeling [40]. Overall, streamlined methods to formulate and solve glycosylation-related mathematical models are awaited.

Besides data collected by individual laboratories, relational databases with systems-level information necessary for mathematical modeling of the glycome are starting to emerge (Table 1). Some of these repositories are a results of efforts by large consortia supported in the USA (Consortium of Functional Glycomics [CFG], www.functionalglycomics.org) [54],

Europe (EuroCarbDB) [55] and Japan (Japan Consortium for Glycobiology and Glycotechnology Database [JCGGDB]). Among these, at the transcript level, the CFG had developed glycogene specific microarrays and deposited related experimental data for two model organisms (human and mouse) under a limited range of experimental conditions [54]. These entries complement more generic transcriptomics data stored in the GEO (Gene Expression Omnibus) database. The JCGGDB also maintains a glycosylation-related disease database, which is an abridged form of the OMIM (Online Mendelian Inheritance in Man) catalog. MALDI-TOF mass spectrometry experiments that profile the N- and O-glycome of cells and tissue by analyzing the distribution of carbohydrates released from their underlying scaffolds are stored at the CFG database. The Unicarb-KB database catalogs intact glycoproteomics analysis mostly from literature [56]. While there is no specific database for the storage of lectin array results, the CFG has developed glycan microarrays that complement these efforts. Here, the binding specificity and affinity of glycan binding proteins (lectins) and antibodies to carbohydrates printed on chips is provided [4,57]. Complementing this, the JCGGDB has mapped the binding specificity of a range of glycan related antibodies (Table 1). The application of computational modeling concepts to quantitatively analyze genomics and proteomics resources stored in these databases is an important future step for our understanding of glycosylation processes.

Conclusions and outlook

The review highlights the manner in which glycans are synthesized and the myriad of control mechanisms regulating this process. The development of computational models to relate data available from different experimental modalities can provide a platform, initially for the qualitative understanding of these glycosylation reaction networks and subsequently for providing quantitative testable hypotheses.

Acknowledgments

SN is supported by NIH/NHLBI grants HL103411 and the Program for Excellence in Glycosciences award HL107146. LKM is supported by the NIH/NIAID grant U01AI111598.

References

1. Varki A. Biological roles of oligosaccharides: All of the theories are correct. Glycobiology. 1993; 3(2):97–130. [PubMed: 8490246]

2. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between o-glcnacylation and phosphorylation: Roles in signaling, transcription, and chronic disease. Annu Rev Biochem. 2011; 80:825–858. [PubMed: 21391816]

3. Cummings RD. The repertoire of glycan determinants in the human glycome. Mol Biosyst. 2009; 5(10):1087–1104. [PubMed: 19756298]

4**. Agrawal P, Kurcon T, Pilobello KT, Rakus JF, Koppolu S, Liu Z, Batista BS, Eng WS, Hsu KL, Liang Y, Mahal LK. Mapping posttranscriptional regulation of the human glycome uncovers microrna defining the glycocode. Proc Natl Acad Sci U S A. 2014; 111(11):4338–4343. Maps the lectin binding profile of the NCI-60 cell lines. Identifies microRNA as a key regulator of cellular glycosylation. [PubMed: 24591635]

5**. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, et al. Precision mapping of the human o-galnac glycoproteome through simplecell technology. EMBO J. 2013; 32(10):1478–1488. Analyzed the spread of O-glycosylation sites using 12 different human cell types that lack the

ability to extend O-GalNAc type carbohydrates (SimpleCells). The study identifies 3000 O-glycosites on 600 glycoproteins. [PubMed: 23584533]

6*. Zielinska DF, Gnad F, Wisniewski JR, Mann M. Precision mapping of an in vivo n-glycoproteome reveals rigid topological and sequence constraints. Cell. 2010; 141(5):897–907. Describes the use of 'filter aided sample preparation' to map 6367 N-glycosylation sites on 2352 proteins in mouse tissue and blood plasma. Besides the consensus N-X-S/T (X≠P) sequon for N-glycosylation, the study also shows rare N-X-C sites for carbohydrate modification. [PubMed: 20510933]

7. Sun S, Shah P, Eshghi ST, Yang W, Trikannad N, Yang S, Chen L, Aiyetan P, Hoti N, Zhang Z, Chan DW, et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of n-linked glycans and glycosite-containing peptides. Nat Biotechnol. 2016; 34(1):84–88. [PubMed: 26571101]

8. Rakus JF, Mahal LK. New technologies for glycomic analysis: Toward a systematic understanding of the glycome. Annual review of analytical chemistry. 2011; 4:367–392.

9. Liu G, Neelamegham S. Integration of systems glycobiology with bioinformatics toolboxes, glycoinformatics resources, and glycoproteomics data. Wiley Interdiscip Rev Syst Biol Med. 2015; 7(4):163–181. [PubMed: 25871730]

10. Cummings RD, Pierce JM. The challenge and promise of glycomics. Chem Biol. 2014; 21(1):1–15. [PubMed: 24439204]

11. Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW. Regulation of glycan structures in animal tissues: Transcript profiling of glycan-related genes. J Biol Chem. 2008; 283(25):17298–17313. [PubMed: 18411279]

12. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013; 152(6): 1237–1251. [PubMed: 23498934]

13. Guo, H.; Pierce, JM. Transcriptional regulation of glycan expression. In: Taniguchi, N.; Endo, T.; Hart, GW.; Seeberger, PH.; Wong, C-H., editors. Glycoscience: Biology and medicine. Springer Japan; Japan: 2015. p. 1568

14. Taniguchi, N.; Honke, K.; Fukuda, M.; Narimatsu, H.; Yamaguchi, Y.; Angata, TE. Handbook of glycosyltransferases and related genes. Springer; Tokyo, Japan: 2014.

15*. Dewal MB, DiChiara AS, Antonopoulos A, Taylor RJ, Harmon CJ, Haslam SM, Dell A, Shoulders MD. Xbp1s links the unfolded protein response to the molecular architecture of mature n-glycans. Chem Biol. 2015; 22(10):1301–1312. An example of how transription factor activity can change the patter of N-linked glycosylation. [PubMed: 26496683]

16. Lau E, Feng Y, Claps G, Fukuda MN, Perlina A, Donn D, Jilaveanu L, Kluger H, Freeze HH, Ronai ZA. The transcription factor atf2 promotes melanoma metastasis by suppressing protein fucosylation. Sci Signal. 2015; 8(406):ra124. [PubMed: 26645581]

17. Vojta A, Samarzija I, Bockor L, Zoldos V. Glyco-genes change expression in cancer through aberrant methylation. Biochim Biophys Acta. 2016

18. Antony P, Rose M, Heidenreich A, Knuchel R, Gaisa NT, Dahl E. Epigenetic inactivation of st6gal1 in human bladder cancer. BMC Cancer. 2014; 14:901. [PubMed: 25465919]

19. Sun Q, Zhang J, Zhou N, Liu X, Shen Y. DNA methylation in cosmc promoter region and aberrantly glycosylated iga1 associated with pediatric iga nephropathy. PLoS One. 2015; 10(2):e0112305. [PubMed: 25647400]

20. Mi R, Song L, Wang Y, Ding X, Zeng J, Lehoux S, Aryal RP, Wang J, Crew VK, van Die I, Chapman AB, et al. Epigenetic silencing of the chaperone cosmc in human leukocytes expressing tn antigen. J Biol Chem. 2012; 287(49):41523–41533. [PubMed: 23035125]

21. Serpa J, Mesquita P, Mendes N, Oliveira C, Almeida R, Santos-Silva F, Reis CA, LePendu J, David L. Expression of lea in gastric cancer cell lines depends on fut3 expression regulated by promoter methylation. Cancer Lett. 2006; 242(2):191–197. [PubMed: 16427187]

22. Fleischer T, Edvardsen H, Solvang HK, Daviaud C, Naume B, Borresen-Dale AL, Kristensen VN, Tost J. Integrated analysis of high-resolution DNA methylation profiles, gene expression, germline genotypes and clinical end points in breast cancer patients. Int J Cancer. 2014; 134(11):2615–2625. [PubMed: 24395279]

23. MacNeil LT, Pons C, Arda HE, Giese GE, Myers CL, Walhout AJ. Transcription factor activity mapping of a tissue-specific gene regulatory network. Cell Syst. 2015; 1(2):152–162. [PubMed: 26430702]
24. Marathe DD, Chandrasekaran EV, Lau JT, Matta KL, Neelamegham S. Systems-level studies of glycosyltransferase gene expression and enzyme activity that are associated with the selectin binding function of human leukocytes. FASEB J. 2008; 22(12):4154–4167. [PubMed: 18716032]
25**. Nairn AV, Aoki K, dela Rosa M, Porterfield M, Lim JM, Kulik M, Pierce JM, Wells L, Dalton S, Tiemeyer M, Moremen KW. Regulation of glycan structures in murine embryonic stem cells: Combined transcript profiling of glycan-related genes and glycan structural analysis. J Biol Chem. 2012; 287(45):37835–37856. Maps glycosylation related transcript levels to corresponding glycan structures. Concludes that glycan structure changes generally (but not always) correlate with alterations in glycogene transcript abundance. [PubMed: 22988249]
26. Anugraham M, Jacob F, Nixdorf S, Everest-Dass AV, Heinzelmann-Schwarz V, Packer NH. Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: Glycan structures reflect gene expression and DNA methylation status. Mol Cell Proteomics. 2014; 13(9): 2213–2232. [PubMed: 24855066]
27. Bartel DP, Chen CZ. Micromanagers of gene expression: The potentially widespread influence of metazoan micrornas. Nat Rev Genet. 2004; 5(5):396–400. [PubMed: 15143321]
28. Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Bluthgen N, Marks DS, van Oudenaarden A. Gene expression. Microrna control of protein expression noise. Science. 2015; 348(6230):128–132. [PubMed: 25838385]
29. Kasper BT, Koppolu S, Mahal LK. Insights into mirna regulation of the human glycome. Biochem Biophys Res Commun. 2014; 445(4):774–779. [PubMed: 24463102]
30. Bernardi C, Soffientini U, Piacente F, Tonetti MG. Effects of micrornas on fucosyltransferase 8 (fut8) expression in hepatocarcinoma cells. PloS one. 2013; 8(10):e76540. [PubMed: 24130780]
31. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009; 324(5924):218–223. [PubMed: 19213877]
32. Kurcon T, Liu Z, Paradkar AV, Vaiana CA, Koppolu S, Agrawal P, Mahal LK. Mirna proxy approach reveals hidden functions of glycosylation. Proc Natl Acad Sci U S A. 2015; 112(23): 7327–7332. [PubMed: 26015571]
33. Vaiana CA, Kurcon T, Mahal LK. Microrna-424 predicts a role for beta-1, 4 branched glycosylation in cell cycle progression. J Biol Chem. 2016; 291(3):1529–1537. [PubMed: 26589799]
34. Patil SA, Chandrasekaran EV, Matta KL, Parikh A, Tzanakakis ES, Neelamegham S. Scaling down the size and increasing the throughput of glycosyltransferase assays: Activity changes on stem cell differentiation. Anal Biochem. 2012; 425(2):135–144. [PubMed: 22449497]
35. Blixt O, Allin K, Bohorov O, Liu X, Andersson-Sand H, Hoffmann J, Razi N. Glycan microarrays for screening sialyltransferase specificities. Glycoconjugate Journal. 2008; 25(1):59–68. [PubMed: 17914671]
36. Houseman BT, Mrksich M. Carbohydrate arrays for the evaluation of protein binding and enzymatic modification. Chem Biol. 2002; 9(4):443–454. [PubMed: 11983333]
37. Northen TR, Lee JC, Hoang L, Raymond J, Hwang DR, Yannone SM, Wong CH, Siuzdak G. A nanostructure-initiator mass spectrometry-based enzyme activity assay. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(10):3678–3683. [PubMed: 18319341]
38. Kudelka MR, Antonopoulos A, Wang Y, Duong DM, Song X, Seyfried NT, Dell A, Haslam SM, Cummings RD, Ju T. Cellular o-glycome reporter/amplification to explore o-glycans of living cells. Nat Methods. 2016; 13(1):81–86. [PubMed: 26619014]
39. Mondal N, Buffone A Jr, Stolfa G, Antonopoulos A, Lau JT, Haslam SM, Dell A, Neelamegham S. St3gal-4 is the primary sialyltransferase regulating the synthesis of e-, p-, and l-selectin ligands on human myeloid leukocytes. Blood. 2015; 125(4):687–696. [PubMed: 25498912]
40. Neelamegham S, Liu G. Systems glycobiology: Biochemical reaction networks regulating glycan structure and function. Glycobiology. 2011; 21(12):1541–1553. [PubMed: 21436236]

41. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, et al. The systems biology markup language (sbml): A medium for representation and exchange of biochemical network models. Bioinformatics. 2003; 19(4):524–531. [PubMed: 12611808]

42. Liu G, Puri A, Neelamegham S. Glycosylation network analysis toolbox: A matlab-based environment for systems glycobiology. Bioinformatics. 2013; 29(3):404–406. [PubMed: 23230149]

43**. Liu G, Neelamegham S. A computational framework for the automated construction of glycosylation reaction networks. PLoS One. 2014; 9(6):e100939. Describes methods to integrate glycan structure information into SBML format pathway models. Also presents algorithms for the automated construction of glycosylation reaction networks using machine readable enzyme definitions. [PubMed: 24978019]

44. Lau KS, Partridge EA, Grigorian A, Silvescu CI, Reinhold VN, Demetriou M, Dennis JW. Complex n-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. Cell. 2007; 129(1):123–134. [PubMed: 17418791]

45. Liu G, Marathe DD, Matta KL, Neelamegham S. Systems-level modeling of cellular glycosylation reaction networks: O-linked glycan formation on natural selectin ligands. Bioinformatics. 2008; 24(23):2740–2747. [PubMed: 18842604]

46. Buffone A Jr, Mondal N, Gupta R, McHugh KP, Lau JT, Neelamegham S. Silencing alpha1, 3-fucosyltransferases in human leukocytes reveals a role for fut9 enzyme during e-selectin-mediated cell adhesion. J Biol Chem. 2013; 288(3):1620–1633. [PubMed: 23192350]

47. Bennun SV, Yarema KJ, Betenbaugh MJ, Krambeck FJ. Integration of the transcriptome and glycome for identification of glycan cell signatures. PLoS Comput Biol. 2013; 9(1):e1002813. [PubMed: 23326219]

48. Puri A, Neelamegham S. Understanding glycomechanics using mathematical modeling: A review of current approaches to simulate cellular glycosylation reaction networks. Ann Biomed Eng. 2012; 40(4):816–827. [PubMed: 22090146]

49. McDonald AG, Tipton KF, Davey GP. A knowledge-based system for display and prediction of o-glycosylation network behaviour in response to enzyme knockouts. PLoS Comput Biol. 2016; 12(4):e1004844. [PubMed: 27054587]

50. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (cazy) in 2013. Nucleic Acids Res. 2014; 42(Database issue):D490–495. [PubMed: 24270786]

51. Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, Kawasaki T, Kanehisa M. Kegg as a glycome informatics resource. Glycobiology. 2006; 16(5):63R–70R.

52. Jedrzejewski PM, del Val IJ, Constantinou A, Dell A, Haslam SM, Polizzi KM, Kontoravdi C. Towards controlling the glycoform: A model framework linking extracellular metabolites to antibody glycosylation. Int J Mol Sci. 2014; 15(3):4492–4522. [PubMed: 24637934]

53. Spahn PN, Hansen AH, Hansen HG, Arnsdorf J, Kildegaard HF, Lewis NE. A markov chain model for n-linked protein glycosylation--towards a low-parameter tool for model-driven glycoengineering. Metab Eng. 2016; 33:52–66. [PubMed: 26537759]

54. Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S, Sasisekharan R. Advancing glycomics: Implementation strategies at the consortium for functional glycomics. Glycobiology. 2006; 16(5):82R–90R.

55. von der Lieth CW, Freire AA, Blank D, Campbell MP, Ceroni A, Damerell DR, Dell A, Dwek RA, Ernst B, Fogh R, Frank M, et al. Eurocarbdb: An open-access platform for glycoinformatics. Glycobiology. 2011; 21(4):493–502. [PubMed: 21106561]

56. Campbell MP, Peterson R, Mariethoz J, Gasteiger E, Akune Y, Aoki-Kinoshita KF, Lisacek F, Packer NH. Unicarbkb: Building a knowledge platform for glycoproteomics. Nucleic Acids Res. 2014; 42(Database issue):D215–221. [PubMed: 24234447]

57. Kletter D, Singh S, Bern M, Haab BB. Global comparisons of lectin-glycan interactions using a database of analyzed glycan array data. Mol Cell Proteomics. 2013; 12(4):1026–1035. [PubMed: 23399549]

HIGHLIGHTS

- Glycosylation is regulated at the transcriptional, translational and post-translational levels
- Advances in sequencing and structural analysis enable data collection at all levels
- Relational databases that store glycosylation-related information are starting to emerge
- Computational models can integrate knowledge across different experimental scales

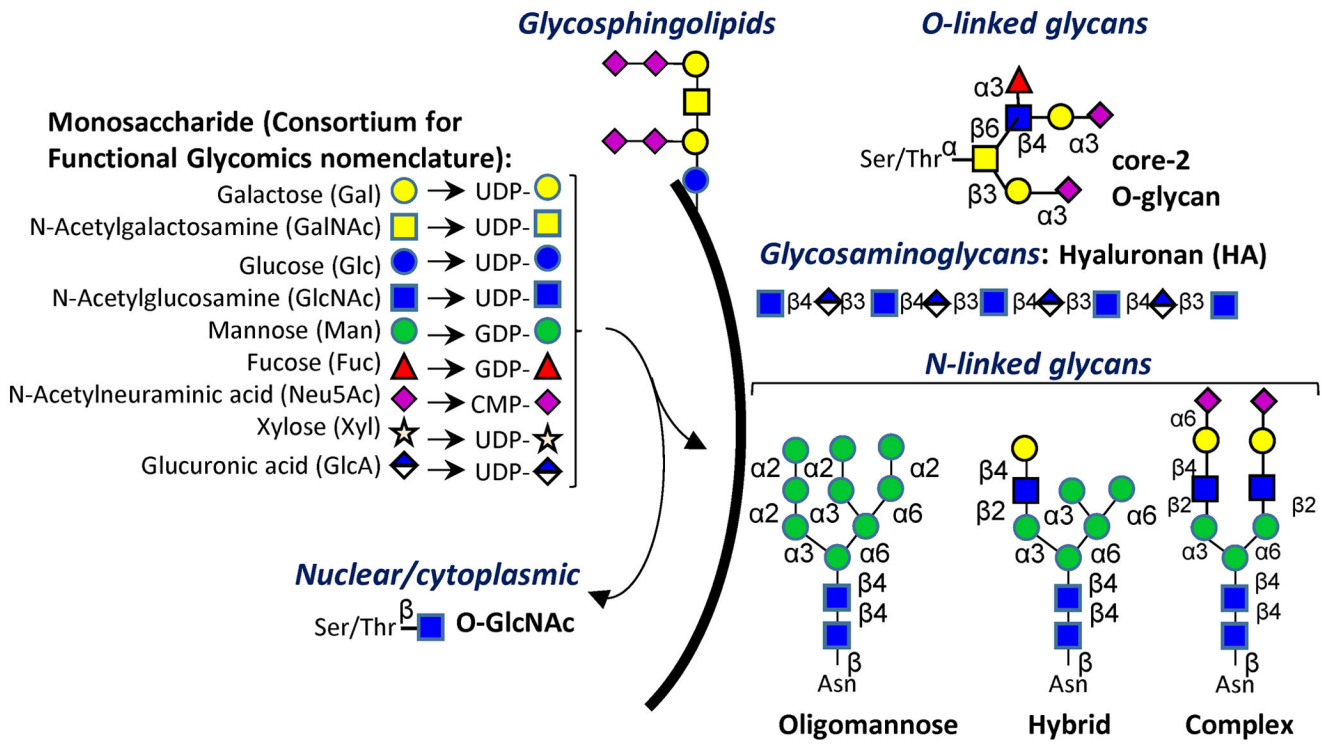


Figure 1. Glycosylation overview

Commonly, nine types of monosaccharides in humans form corresponding sugar-nucleotides (UDP-Gal etc.). A variety of glycoenzymes (particularly the glycosyltransferases and glycosidases) enable the transfer of these monosaccharides to lipid and protein scaffolds. Additional monosaccharides can form, e.g. Iduronoic acid (IdoA), by the epimerization of Glucuronic acid (GlcA) on glycosaminoglycan chains. Carbohydrate structures thus formed decorate the cell surface, and they are also found in intra-cellular compartments. Most glycan families have a limited number of core-structures that are elaborated by repeat extensions (like N-Acetyllactosamine repeats) and terminal modifications.

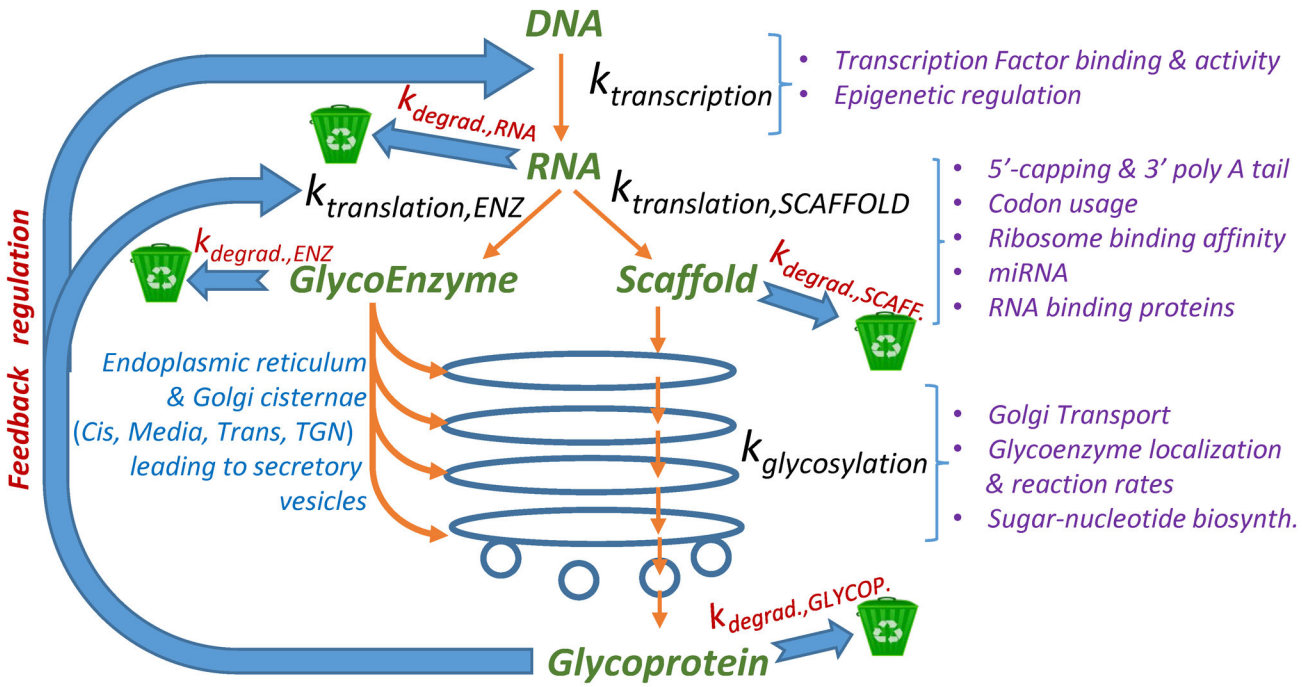


Figure 2. Regulation of glycosylation

The central dogma of molecular biology states the passage of information from DNA to RNA to protein in all living organisms. These proteins can be further glycosylated by the glycoenzymes, mainly in the endoplasmic reticulum and Golgi compartments. RNA and protein degradation rates, shown using recycle bins, are important checkpoints that control cell surface carbohydrate structures. Additional parameters that track the process of information transfer from genes to RNA to proteins (including glycoenzymes) to glycoproteins are shown using purple text. Glycoproteins may provide feedback and feedforward control to regulate cellular transcription and translation. The parameter '*k*' (e.g. $k_{transcription}$ and $k_{degrad.}$) is used to denote lumped rate constants for individual steps.

Table I

Databases for the Glycosciences *

Level	Description and unique feature	Website
Glycan entries	GlyTouCan	glytoucan.org
	UnicarbKB	unicarbkb.org
Enzyme data	CAZY (Carbohydrate-Active Enzyme)	www.cazy.org
	IUBMB ExploreEnz database	enzyme-database.org
	BRENDA (Comoprehensive enzyme info. system)	brenda-enzymes.org
Transcript	Glycogene microarray data - for human/mouse samples (CFG)	www.functionalglycomics.org
	Gene Expression Omnibus (GEO)- functional genomics repository	www.ncbi.nlm.nih.gov/geo
Glycomics/Glycoproteomics	GlycoBase	glycobase.nibrt.ie
	UniCarb-KB	www.unicarbkb.org
	CFG MALDI-TOF/TOF datasets	www.functionalglycomics.org
Glycan binding data	CFG glycan array entries	www.functionalglycomics.org
	LfDB (Lectin Frontier Database) & GlycoEpitope	jcggdb.jp/index_en.html
Pathway	KEGG GLYCAN	www.genome.jp/kegg/glycan

* This table only lists relational databases that would be useful for computational modeling of cellular glycosylation biosynthetic pathways. For a more comprehensive list of all glycan-related databases, please see ref. [9].